DOCUMENT RESUME

ED 242 779 TM 840 197

AUTHOR Ford, Marge L.; Stevens, Patricia C.
TITLE Institutional Research Problems Related to Large Databases at Multicampus Universities.
PUB DATE Nov 83
NOTE 10p.; Paper presented at the Annual Meeting of the Mid-America Regional Institutional Research Conference (Kansas City, MO, November 7-8, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Computers; Computer Storage Devices; *Databases; Higher Education; Information Storage; *Management Information Systems; *Multicampus Colleges; *Research Problems
IDENTIFIERS Integrated Databases; *Transactional Data Systems

ABSTRACT

Large central databases which support multicampus users require thorough planning and coordination, special communication with both users and input sources, and organized retrieval systems. In the developmental process of identifying variables that should be in the system and designating the source of that data, old habits of multiple official paper records must be altered. As data from feeder systems flow into the central database, ownership of the data and methods and levels of access become critical issues. However, there must be clearly specified the protocol for entry and updating of data and for verification of its correctness and timeliness. Frozen files which follow through edit procedures and pre-specified multi-purpose reports enhance the functional operation of the systems. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED242779

INSTITUTIONAL RESEARCH PROBLEMS RELATED TO LARGE DATABASES AT MULTICAMPUS UNIVERSITIES

Marge L. Ford, Ed. D.
Director Institutional Research

Patricia C. Stevens, Ph.D.
Institutional Research Associate

University of Missouri-Kansas City

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

P.C. Stevens &
M.L. Ford

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 810 157

INTRODUCTION

In almost all institutions of higher education today, fairly sophisticated information systems exist, many of which are computerized to support the operational and decision-making processes of the institution. Information systems, which are a result of transactional systems, routinely operate both manually and with considerable computer support. (Parker, 1978.) A transactional system is an established set of procedures for gathering, filing and recalling data which is mostly in disaggregrated form and format. (Campbell, 1977.) The data, which begins as part of a specific file within a major single line office such as the registrar or budget office, is used in the day-to-day operation to fulfill the functions which are the role of that office. The determination of what data is to be collected and how it is to be placed and maintained in the file is within the domain of the office.

Information systems contain the data in the various files of the institutions' offices and it is from these files that analyses are made to provide information which can be used for decision making.

In a large university, such as the University of Missouri, transactional data is a part of a computerized central data base with the campus office controlling the data input. The data output for informational and transactional requirements is produced centrally. There are numerous problems which arise for the day-to-day user, as well as those of us in institutional research who must provide an analysis of the information found in the files of the system. This paper discusses some of the problems which are related to large central integrated data bases within a large university system.

Satisfying the Requirements of Central Data Base Systems

In a manual system, which services one single office on a campus, it is fairly easy to determine what data will be collected and maintained within the file structure of that office. In most universities and colleges, a single file for functions of one office has developed independently of files in another office, even though both may have many common data elements. Determining what would be kept in the records of the file is usually the function of the personnel who use the file within the office. As long as each office had its own data to support the office's functions and for the most part could operate independently from the other offices on campus, there were few problems. At least in terms of the transactions, there were few problems.

However, under such circumstances, we institutional researchers find a great deal of inconsistency in the data in the various files which we use for analysis, even when the independent files are controlled by a single office. When we have a report which requires the analysis of a single office's data, such as admissions, we relegate the task to the office which has the data. When we do this, we must remind ourselves that data analysis is not the primary function of that office and will not be approached with the same enthusiasm or expertise of the

institutional researcher. The report may not be completed with the same level of understanding of what is required. Also, since we are not owners of the data, we may not be able to verify the data reported.

However, a more critical problem is present when departments have variant definitions for the data elements and/or the analysis units. Thus it is possible for two departments to respond to a survey on behalf of the institution and to provide different data. This leads to confusion as well as loss of creditability. As Institutional Researchers, we know that external reports frequently provide definitions of terms and variables considered, and these often differ between reports. Individual departments seldom know that this is true and may never consider the existence of variant definitions, but rather use what is convenient.

Until recently, at the University of Missouri, most data bases were established for use by a specific department as described previously. These data bases were essentially accessed and controlled by a single department. The University has moved from a manual single file system to a large integrated data base system with multiple users. Centrally controlled, batch oriented systems are being replaced by terminal based transactions entered by the user in the user's office.

With the integrated data bases, the data determined and collected for entry into the file must meet the demands of several users. Theoretically, such a system puts the transactional data in the hands of the user, with all users working with a commonly defined base.

Integrated data bases must have data definitions which are the same for all users and which can be understood by the users in all units. It is difficult in the development of the definitions of the data elements to ascertain the needs of all users. Generally, the systems developers look first to the primary users. In large systems, not only the users within the specific functions must be considered, but the unique requirements of the campus must be addressed. Specificity of definition is necessary to the system so that interpretation of the element definitions will be the same across all units. Treating unique data requirements within the realm of the data definitions may be difficult. The development of a system with the adequate elements and common definitions which meet the requirements of transactional users and informational analysis is a multi-step process. The most obvious point of coordination is at the interface between the systems developers in the central offices and the various campus users. Generally, this is accomplished by a task force with representation from the various user groups. However, the central effort of the systems designer is only as effective as the input received. At the campus level, someone must identify the pieces of information needed by each relevant department. This includes going beyond obvious needs, such as a specific report or function, to determine other data required during the year, or the data needed which may be unique to a specific unit or unit functions.

For example, if a payroll/personnel system is developed based on payroll requirements, personnel information may be problematic. If appointments exist for a person who might have zero pay, but an FTE appointment of anywhere from .05 to 1.0 FTE, it would become an edit or could be aborted from the system. We have a number of faculty on our campus which fit this description. These faculty have appointments and serve the University; they need to be a part of the personnel records. They are eligible for promotion, yet they do not fit the parameters of the system. Fortunately, this was resolved, but the problem emerged from the lack of proper input during the development period of the system. When users, particularly those who have begun to use a computerized system, do not have their unique transactional data processing requirements met, they tend to duplicate the records in their offices. Also, when the integrated data bases are installed, access to data may no longer be readily available as it was previously, particularly as it was with a manual file system.

In order for data base systems to become an advantage to the transactional user, such that they will rely on it and give up their duplicate systems, the data base must do more than just contain the proper elements; there must be adequate levels of data reliability and data access. Reliability implies correctness of data, where correct means both the lack of mistakes and also currency, based on timely data entry. Additionally, reliability requires that the data base is complete, including all the requisite elements and all the relevant cases.

Access to the data must be both timely and easy. Day-to-day information for the transaction user may be through interactive terminals or by hard copy printouts. Direct element users probably should have a terminal inquiry. For the person seeking specific information to answer questions or concerns, it is important that the information be provided in a timely manner. Generally, these answers come through the office of institutional research which has access to the various reports produced centrally from the system.

One of the features of a large central data base, as was mentioned earlier, is the consistent data definitions and of course, the centrally compiled variables which can be analyzed as part of the information system. However, the more complex the system, the more complicated the development of the output containing the required information. Programming in some instances may not be feasible. As a result, multipurpose reports may be more desirable.

First, such a report may serve many users needs. Instead of developing and running many separate programs with a limited number of items for single users, the same report may be expanded and with certain select options serve the needs of many users. Secondly, due to reporting demands of external agencies, many of the institution's reports have been developed based on agency definitions and requirements,

and these are of value in the development of institutional information reports. Some specific informational requirements can come from further analysis using the output, or using the report program and having it modified to do further analysis. It should be noted that the latter can occur only if the report has been developed in such a manner that it can be modified for additional purposes, generally using options, and if the additional modification does not change the original purpose for which it was designed.

Feeder Systems and Ownership

Until recently in the University of Missouri system, most data bases were established for use by a specific department and these data bases were essentially accessed and controlled by a single department. The establishment of data bases to be accessed by multiple users with varying requirements raises an important issue as to ownership.

There are many types of information within the University with substantial centrally located data bases containing information on students, employees, library holdings, finances, etc. and there are many other smaller data bases which are in use by individual departments or divisions.

The University has adopted the term 'custodian' for multiple use data bases where ownership can not be relegated to a single department. Under the most current definitions of the University, such data bases are considered to be a part of the computing center. Specifications for the requirements of the data base and input and output control are the function of the computer center. Although the campus has input in the development specifications, the system controls, operations, devices or data entry and reports of the system are products of the computer center. Each data base has a data base custodian at the campus level who is responsible for the integrity of the data. The integrity is defined as the quality, security and confidentiality of the data contained within the data base and also all the data extracted from the data base in printed format, magnetically encoded and/or terminal accessed.

The data base custodian for each of the data systems, such as the student information system or personnel system, has the control at the campus level for determining the points of data entry, particularly on-line entry, authorizes changes, and develops and oversees procedures for updating the file. The data base custodian may be a person in the Registrar's office who would oversee the student information systems. Similarly, there may be a data base custodian in the Personnel Office to oversee the personnel file. Generally, the data bases encompass information absolutely necessary to the operation of other units. For example, payroll office staff are heavily dependent upon data in the personnel file and must have access to the data in the file. Custodians must recognize the responsibility for meeting these various user needs.

Institutional researchers can be helpful to the custodians and serve as a liaison between the various custodians, as well as other users needing data from the systems. Informational needs can be interpreted to the custodians and the institutional researcher can oversee the data analysis to assure the security and privacy requirements of the university are met. These requirements and limitations of the data base can also be transmitted to the users requesting the data.

Data entry is generally the responsibility of the transactional users. For example, the personnel office staff enter employee data in the system. The data which comes to the personnel office, however, may come from many various units across the campus. There must be coordination so that the paper flow which provides the data entry is received in a timely fashion and that the coding meets the data definition specification. Correct data is extremely important to the system.

Verification of data is a means for assuring the correctness of the data base. It is important not only from an operational point of view, but it is extremely important when analyses are to be done with the data base. Particularly when a file is to become an official file or report file.

Verification can occur upon entry of the data into the system or at a later date in the preparation of some report or as a part of the general maintenance procedures. Some obvious edit procedures include a check for missing data or nonsense data. There are programs which can check the existence of data for a designated variable to determine if the data conforms to legitimate codes specified for the variable. However, there is no edit program which could find incorrect data such as the entry of F for sex code for a male employee. Inaccuracies in the system can be checked by taking a sample from the total data base and checking the data element entries against a hard copy file. It is important when the data base is developed that, as much as possible, edit samples be used to determine the validity of the data base. The data in the system is no better than the input and there will always be the possibility of some invalid data.

Data base file systems are fluid; that is, the information entered into the data base or the file changes from day to day. Because of report and audit requirements, most institutions develop an official or frozen file. A snapshot is taken at a certain point in time as designated by agency reporting requirements or by campus consensus. Generally, by a certain date, all pertinent data will be entered into the file as specified to meet the various reporting and audit requirements. Often these files are used for the generation of internal reports as well.

Report Development

Generally speaking, systems staff or the institutional research officer for the system serves as the coordinator for the development of reports which come from the data bases. They are much more aware of the general requirements of the university. Also, they often have responsibility for the coordination of the completion of various external reports. Campus staff can bring to their attention questions about reports which are needed and which could be developed as a multi-campus effort. The institutional researcher at the campus level can serve as the liaison between the various units on campus and the central offices to assure that the campus concerns are addressed.

Often when there is no institutional research office at the campus level, the central administrative institutional research office must work with the custodians at each campus, as well as stay abreast of the various informational requirements at the campus level. In such cases, information becomes more centrally oriented; often, information that is specific or unique to the individual campus needs for planning or decision making is not made available. Report generation under this arrangement often becomes a secondary function of the office of the custodian of the data base.

The institutional research office, at the campus level is the conduit for campus level research and for data necessary to explore and support administrative decisions. Such information need not be generalized across campuses and may be required at many various times throughout the year. Some of the data will be on existing computer reports, particularly where reports have been designed as multipurpose. In other cases, the data can be derived from existing computer printouts or central administration reports, but this may require laborious, time-consuming, manual manipulation. An alternative is to maintain a duplicate data base for report generation specific to campus needs.

Data requests should be carefully scrutinized to determine if a request will be recurring from semester to semester or if it is a one-time request. Before developing a program for a specific data request, care should be taken to consider the long range value of such an effort. It may be just as cost effective to make one-time calculations manually.

Most institutions have not made provisions for providing information which can be easily retrieved from data bases for local analysis. Complexity of the various data base systems and the fact that often the files in a systems do not speak to one another complicates the situation further. As pointed out earlier, such data bases are built to meet certain transactional functions.

User friendly software, such as SAS or SPSS which we might use in institutional research can not be used against a duplicate tape of the

data base without extensive reformating in most cases. Campuses must weigh the results of such a programming endeavor against its cost to provide expediency of data analysis. Often, the campus must continue to rely on programming at the central level and reports of a more general nature.

SUMMARY

University systems, such as the University of Missouri which serves several campuses, are moving from a manual, single purpose file informational system to an integrated data base system. The data collected and maintained in the system is no longer a single office function, but instead must meet the demands of many users. Files are developed so that they can be merged. Central data base systems provide for common data definitions with the input and the output designs being coordinated by a central computing center. Campus input as to the specifications of transactional and informational data needs are imperative to a good system. Generally, with central data base systems, there also exist campus duplicate systems. For the user to depend upon the central system totally, there must be reliability of data and easy access. It must provide the user easy input and output to meet specific day-to-day functions, as well as provide for data analysis to meet various reporting requirements. If this provision is not made, the user will develop an independent system to meet those requirements in a timely fashion.

The needs of a transactional system of an integrated data base system predetermine the elements for informational retrieval and analysis. Each campus has a data base custodian who is responsible for the particular file system. The campus institutional researcher can assist both the user and custodian in determining analysis and reporting requirements.

Generally speaking, institutional researchers at the central office serve as the coordinators for the development of reports from the data bases. Often specific data needs can be taken from centrally developed multipurpose reports. Generally, campuses do not have the luxury of expedient data analysis from various files without extensive programming or reformating of a file to use with data analysis programs.

All in all, the ability to meet the informational requests of most users is present in a university system such as the University of Missouri. It just takes some know-how to find the specific transactional document which provides the requisite data or to access the resources available to develop a report generation file which can support user friendly software.

References:

Campbell, Bonita J. , Understanding Information Systems. Cambridge, Mass., Winthrop Publishers, Inc. 1977

Parker, William C., The Management Information System in Higher Education: A Duper Look at the (MIS) and Its Development and Implementation: Washington, D.C., U. S. Office of Education, 1977